

An overview of procurement integrity and introduction to opentender.eu

## Procurement data: sources, possible errors, and examples of data availability

Péter Horn

Second annual summer school of the R2G4P, 8 June 2022

Implemented by:



The R2G4P project, coordinated by the Center for the Study of Democracy, Bulgaria benefits from a € 1.5 million grant from Iceland, Liechtenstein and Norway through the EEA and Norway Grants Fund for Regional Cooperation. The aim of the project is to implement shared anti-corruption and good governance solutions in Southeast Europe through innovative practices and public-private partnerships.

#### **Presentation overview**

- The goals of a procurement database
- Relevant datatypes
- Key aspects of procurement data (scope, depth, quality, access)
- Examples from partner countries
- Errors in the data

## Goals & Objectives of procurement databases

#### Main objective:

- Goal: To create a comprehensive public procurement dataset, which makes the evaluation of the integrity of countries' procurement systems possible.
- Create objective indicators to measure procurement integrity/corruption risk
- This requires high quality administrative data on:
  - 1. public procurement tenders and contracts,
  - 2. bidding companies,
  - 3. awarding public organizations and
  - 4. political office holders.

#### DIGIWHIST:

- Large-scale EU-funded research project which simultaneously aims to increase trust in governments and improve the efficiency of public spending across Europe.
- Supports corruption measurement by organizing and linking the four complex datasets.
- Its data template also serves the basis for collecting and republishing publicly available and sufficiently well-structured databases pertaining to corruption measurement in Europe.

## Data types I.

- 1. Public procurement data (contract or item level) Mostly available
  - 1. Call for tender related information: procedure type, product code, bidding period length, bidder limitation, estimated value, type of the contract, documentation fee, buyer, award criteria.
  - 2. Contract award related information: number of bids received, bidder and winner company related information (bid prices, location), final contract value, award signature date.

#### 2. Company data - Partially available

- 1. **Registry information:** company name, location, legal form, date of incorporation, number of employees etc.
- 2. Financial information: annual turnover, profit rate, return on assets, material costs, personnel costs, taxes, EBITDA.
- 3. **Ownership information:** number of recorded shareholders, shareholder's name, shareholder's type (legal entity, individual etc.), shareholder's location, shareholder's direct and total shares.
- 4. **Manager information:** number of directors, name of company directors, position of company directors, appointment and resignation date of directors, gender, date of birth, shareholder status.

### Data types II.

- 3. Public organization data Partially available
  - **3. Registry data**: name, ID, location, activity type, contact information.
  - 4. Budget data: annual budget figures, currency, classification of the budget item (IFRS).
- 4. Public officials' data Mostly unavailable
  - 1. Name, contracting authority, position, start and end date, political affiliation.

## Key aspects of procurement data

- 1. Scope: The range of transactions the publicly available procurement data covers
  - E.g., publishing threshold
- 2. Depth: Amount of information disclosed for each contracts/tenders
  - Depth of information within each data types (e.g., does budget data available for public organizations or only registry data?)
- 3. Quality: Reliability of the data, share of missing information
- 4. Access: How easy is it to obtain the procurement data?
  - Is there an API or the website has to be scraped?

## 1. Data Scope I. – Reporting thresholds

**Reporting thresholds**: National contract value thresholds for mandatory publication of tenders on national or EU-wide portals

- Procurements over the threshold also have to comply with stricter rules
  - such as minimum length of advertisement period, or publication of scoring criteria.
  - Hence, lower threshold leads to more transparency.
  - Tenders under the threshold are significantly more likely to have restricted types (e.g., direct awards, negotiated tenders)

Reporting thresholds can have different meaning across countries and across time (e.g., in Turkey several public bodies are exempt from the threshold)

#### Scope of public procurement databases

Minimum contract value for publishing supplies and services contracts (EUR, 2015)



#### 1. Data Scope II. – Number of tenders processed by DIGIWHIST

- Result of the variation in thresholds is that publicly available data quantity largely differs
- More data leads to better/less biased analysis

Number of contracts collected by DIGIWHIST per country TED + National data, 2006(2007) - 2020



### 2. Data depth I. - Tender cycle

The tender cycle consists of:



## 2. Data depth II. – Tender cycle coverage

Problems with tender cycle coverage in Europe:

- 1. Most of the procurement systems in Europe only cover the tendering phase and the awarding phase.
  - Only a handful of EU countries' procurement system disclose information on implementation
  - No information on implementation can give a false picture of the procurement
    - The project can go over budget, or it can be poorly implemented
- 2. The depth of information within a cycle can vary greatly across countries, due to different (and often changing) legislature
  - 1. E.g., the UK does not collect bidder number information significantly reducing data usability
- 3. In many public procurement data systems, modifications and failed tenders are not adequately logged
  - There is no data point indicating tender failure, making failed tenders look like tenders with incomplete information

#### Coverage of the full tender cycle, 2016



Available

Not available

## 2. Data depth III. – Organizational IDs

- Many countries only publish the name and location of organization without any unique identifiers
- Organizational IDs for buyers and suppliers are important, to track organizations over time
  - E.g., how different firms perform across different governments
- Without IDs only name-location pairs can be used to identify different buyers and suppliers
  - These can change over time and prone to grammatical errors (typos)

**Coverage of organizational IDs, 2016** 



## 2. Data depth IV. – Minimum data scope

Minimum required information for comprehensive corruption risk assessment

Variable group	Variable
Buyer	Buyer's name, Buyer's unique ID, Buyer's address
Bidder/bids	Bidder's name, Bidder's unique ID/tax ID, Bidder's address, Number of bids submitted, Number of bids excluded, Bid price, Exact time of bid submission, Bid type (winner/loser bid), Beneficial owners
Tender/contract	Procedure type, Framework agreement, Estimated price, Procurement type (service, supply, work), CPV codes, NUTS codes, Status (cancelled, pending etc.)
Dates	Call for tender publication date, Bid submission deadline, Contract start and end dates, Publication date of contract award, Date of contract completion
Subcontracting	Subcontractor's name and unique ID, Subcontractor's share
Consortium	Consortium members' name and unique ID, Consortium member's unique ID
Contract performance	Contract performance end date, Was performed according to contract, Explanation in case of deferring from contract, Information on contract modification, Information on performance quality

## 3. Data quality – Share of missing information

- In some countries even legally required administrative information is missing from tender announcements
  - Such as buyer name, tender price, bidder information etc.
- Data quality is low throughout Europe with 15% of mandatory fields are missing on average
- Eastern-European countries pp system fare much better than more developed nations'

**Extent of missing information** EU-wide TED data between 2009-2015



## 4. Data accessibility I. – Extraction method

#### Goal:

To create structured database from non-structured/semistructured (text, html, pdf) data

#### Method:

- Prerequisite (min. requirement) is machine readability. (HTML, readable PDF)
- Web crawling/scraping → collecting the data from the webpage (Java, Python, R)
- 2. Database creation (JSON, NOSQL, MongoDB)
- Parsing → automatic text extraction to create data from text (Human assisted) data correction / cleaning, imputation
- 4. Testing data quality (manual/automatic)

Data analysis and indicator creation



## 4. Data accessibility II. - Machine readability I.

Data can be obtained in a:

- 1. Structured format
  - Whole dataset can be downloaded into an excel/json file
- 2. Semi-structured format (Semi machine-readable)
  - Information is available in a html format, can be scraped and parsed
- 3. Not fully machine readable
  - Part of the data only accessible by manual cleaning (e.g., scanned pdfs)
- 4. No public database

## 4. Data accessibility II. - Machine readability II.

## Machine readability & automatization can be problematic:

- Only in three countries and the EU-wide TED's public procurement data is machine readable in a structured format
- In 26 countries, data is only semi machinereadable,
- In 5 countries it is not machine-readable or has no public data av.
- These barriers prevent researchers and NGO's to efficiently analyze the region's public procurement systems

#### Machine readability of pp databases in European



## 4. Data accessibility III. – Usual data sources

#### Sources:

- Open accessibility requires data sources to be publicly available such as:
  - National procurement websites (etendering)
  - EU's Tender Electronic Daily (TED)
  - Public organizations' registry and budget data

## DIGIWHIST data on opentender.eu:

- Over 40 million public contracts from 32 countries (more to come) in a standardized format
  - JSON, CSV, NDJSON
- Over 5 million government suppliers and 1 million public organizations

More on this in the last section...



## Examples from partner countries I. – North Macedonia

#### The good:

- The Electronic System for PP (ESPP) was set up and running in 2006 (Longest running system in the region)
- Complete tender documentation is required to publish new tender notice
- The publication includes information on all phases of procurement from planning to contract implementation (this is hard to get).
- Compared to the other countries in the Western Balkans, North Macedonia performs highest on accessibility and usability of standard data fields

- Most of the organization ID-s are missing
  - ▶ 71% of buyers and 99% for suppliers
- Adding full data download (or API) could further improve accessibility



## Examples from partner countries II. – Serbia

#### The good:

- New procurement website since 2020.
- Certain information on tenders and contracts is available in a semi-machine readable format (html) in both the old and new websites
- The new portal gives an opportunity to download data in XLSX, XML or Json formats
- Organization IDs and tender number is available, which allows matching databases

- Only new tenders are recorded in the new website
- Some of the attachments are non-machine readable (e.g., scanned pdf-s)
- Downloadable datasets only include buyer name, tender ID and title, date of publication, type of announcement
- Every other information must be traced back from the website



# Examples from partner countries III. – Bosnia and Herzegovina

#### The good:

- Certain information on tenders and contracts is available in a semi-machine readable format (html)
- All the information on buyers and bidders (name, ID, address, agency type) is published in standardized (and readable) pdf format, the same is true for contract details, number of bids, eligibility details and deadlines.

- More than ¼ of the pp data is missing
- Most of the details only available in PDF (not html)
- The source lacks pre-tender information such as procurement plans, as well as details on supplier's performance or contract completion
- The details provided in PDFs differs by tender



## Examples from partner countries III. – Montenegro

- Certain information on tenders and contracts is available in a semi-machine readable format (html)
  - > Type of procurement, price details such as estimated value and currency
- Limited information can be exported in CSV, XLSX, XLM and PDF format

- ▶ More than <sup>1</sup>⁄<sub>3</sub> of the pp data is missing
- Most of the information is provided in many separate world/pdf files.
  - Many files are scanned, and badly structured lengthy documents
  - Even within one tender the types of the documents may differ
- Bidder and buyer IDs are always absent
- Exportable information is limited and only one page can be downloaded at once



### Possible errors in the data I. – Common errors

#### Common errors:

- Missing attributes  $\rightarrow$  No information provided (not necessary an error).
- Lexical error → The value is provided is not consistent with the column name (e.g., country id column shows currency id)
- Irregularity error  $\rightarrow$  E.g., the unit of measurement differs from the other observations'
- Formatting error  $\rightarrow$  E.g., date is in different time format leading to errors when data is loaded
- Duplication error  $\rightarrow$  There are duplicate observations in the data; each variable is the same
- ► Contradiction error → Two columns measuring (almost) the same thing show different values for the same observation.
- ► Outlier → Given variable for a given observation is significantly different from the others (not necessary an error, but usually should be delt with)

## Possible errors in the data II. – Example of lexical error

- Figure shows relative price distribution of tenders below and above the EU-threshold
- The distribution has two "peaks" because in some cases prices were recorded with VAT even though a net value should have been recorded



## Data wrangling good practice

No dataset is unique to a different set of errors; hence it is always important to:

- Explore the data before deeper analysis (check column values, distribution for numeric columns, averages, etc.).
- Only use columns that are sufficiently "clean" and not much is missing (~15%)
- Cross-check/sanity-check every result
- Use alternative qualitative methods such as expert interviews
- Procurement data is always just an approximation of reality
  - With many information missing or contradicting we cannot see the whole picture, hence all results should be treated in their proper place

## Thank you!

## Questions?

nfL

Ц Ц

## Sources

- Mara Mendes, Mihály Fazekas (2015): DIGIWHIST Recommendations for the Implementation of Open Public Procurement Data - An Implementer's Guide
- Mihály Fazekas, Luciana Cingolani, Bence Tóth (2016): A comprehensive review of objective corruption proxies in public procurement: risky actors, transactions, and vehicles of rent extraction, Working Paper series: GTI-WP/2016:03, link
- Mihály Fazekas (2021): Using Macro Analytics (Big Data Analytics) to Monitor Fraud and Corruption in Public Procurement In: Using Data Analytics To Combat Fraud And Corruption In Public Procurement In Healthcare In Eastern Europe, Ceeli Institute, <u>link</u>
- Mihály Fazekas, Ágnes Czibik (2021): Measuring regional quality of government: the public spending quality index based on government contracting data, Regional Studies, DOI: 10.1080/00343404.2021.1902975
- Mihály Fazekas (2021): Corruption Risks in Public Procurement in the Western Balkans and Turkey (Draft)
- Nicolas Rangeon, Anderson Prewitt (s.a.): Identify Different Types of Errors, <u>link</u>

Iceland Liechtenstein Norway Norway grants grants

#### An overview of procurement integrity and introduction to opentender.eu

## Introduction to **Opentender** (back-up slides)

Péter Horn

Second annual summer school of the R2G4P, 8 June 2022

Implemented by:



The R2G4P project, coordinated by the Center for the Study of Democracy, Bulgaria benefits from a € 1.5 million grant from Iceland, Liechtenstein and Norway through the EEA and Norway Grants Fund for Regional Cooperation. The aim of the project is to implement shared anti-corruption and good governance solutions in Southeast Europe through innovative practices and public-private partnerships.

## Starting page, country sellection

You can view the data on individual national portals by selecting a country below or explore all available data.



Leaflet | Map tiles by Carto, under CC BY 3.0. Data by OpenStreetMap, under ODbL

			INF. OF Tenders
Austria	95,156	Italy	14.0 Million
Belgium	96,256	Latvia	137,110
Bulgaria	162,316	Lithuania	189,537
Croatia	241,582	Luxembourg	8,176
Cyprus	8,944	Malta	7,718
Czech Republic	253,467	Netherlands	120,604
Denmark	47,356	Norway	254,763
Estonia	104,725	Poland	2.9 Million
EU Institutions	178,243	Portugal	1.2 Million
Finland	59,632	Romania	20.3 Million
France	2.7 Million	Slovakia	303,342
Georgia	329,150	Slovenia	126,962
Germany	483,655	Spain	1.4 Million
Greece	57,906	Sweden	116,284
Hungary	197,417	Switzerland	103,233
Iceland	2,600	United Kingdom	458,749
Ireland	125,296		

## Main country page



Data

HOME DASHBOARDS SEARCH DOWNLOAD ABOUT DONATE



### Country level procurrement market analysis

#### CROATIA

HOME DASHBOARDS SEARCH DOWNLOAD ABOUT DONATE

MARKET ANALYSIS ADMINISTRATIVE CAPACITY TRANSPARENCY INTEGRITY





## **Country level integrity indicator analysis**

#### opentender CROATIA

HOME DASHBOARDS SEARCH DOWNLOAD ABOUT DONATE

MARKET ANALYSIS ADMINISTRATIVE CAPACITY TRANSPARENCY INTEGRITY





Score Range

20



## Search for tenders

CROATIA	HOME DASHBOARDS SEARCH DOWNLOAD ABOUT DONATE	
	TENDER COMPANY AUTHORITY	
	Search Tender	
		Filter Click to Select •
	Buyer: Name Supplier: Name Tender: Title	×
	Score: Integrity     Score: Administrative Capacity     Dates: Award Decision Year     Score: Transpare       0     100     100     2011     2021     0	xncy X
	0 - 100 0 - 100 2009 - 2018	0 - 100

Visualization Click to Select •

Columns Click to Select •

#### Tenders: 241,582

ender Link	Title	Buyer	Supplier	Bid Price
	Medicinska i nemedicinska oprema za dnevne bolnice Opće županijske bolnice Vukovar i bolnice hrvatskih veterana (EU projekt)	n Opća županijska bolnica Vukovar i bolnica hrvatskih veterana	Lot 9 The B.Braun Adria d.o.o. Lot 4 The Boneno d.o.o. Lot 5 The Doctum d.o.o. Lot 3 The Elektroničar d.o.o Lot 1 The Končar elektronika i informatika d.d. Lot 10 The Mel - Medikal d.o.o. Lot 12 The THNOMEDIKA d.o.o.	11 Entries
	Usluga zaštite osoba i imovine za razdoblje od dvije godine	🏦 Rijeka sport d.o.o.		

### View specific tenders

Medicinska i nemedicinska oprema za dnevne bolnice Opće županijske bolnice Vukovar i bolnice hrvatskih veterana (EU projekt)

tenders.exposed	
Find Freedom of Information	
requests at	
Imamo pravo znati	

#### Medical equipments

Indicators ^

Administrative Capacity Indicator



#### Description ^

Navedeno u DON-u.

#### Buyer ^

 Opća županijska bolnica Vukovar i bolnica hrvatskih veterana Županijska 35 32000, Vukovar NUTS HR04C Croatia

Appeal Body Name

Državna komisija za kontrolu postupaka javne nabave Koturaška cesta 43/IVZagreb10000+385 14559930dkom@dkom.hr+385 14559933www.dkom.hr

#### Tender Information ^

Prices		Types	
Estimated Price	€ 240.807	Supply Type	Supplie
	(national) € 240,807	Procedure Type	Oper
	(VAT) 25%	Selection Method	Mea
Final Price	€ 184,972	Eligible Bid Languages	HF
	(national) € 184,972		
	(VAT) 25%		
Dates			
Bid Deadline	20/12/2018 10:00		
Sectors			
Lot 1 - Računalna i elektronička oprema			
🖁 Končar elektronika i informatika d.d	., Zagreb		Price € 35,706 (national) € 35,706
Lot 2 ~			
Oprema za kineziterapiju			
🖣 Elektroničar d.o.o, Zagreb			Price € 6,349 (national) € 6,349

#### Search for contractors

#### opentender CROATIA

HOME DASHBOARDS SEARCH DOWNLOAD ABOUT DONATE

TENDER COMPANY AUTHORITY

#### Search Company

Filter Click to Select\*

Company: Name	Company: City
Company: Name	Company: City
Zajednica ponuditelja GRENOBLOISE D'ELECTRONIQUE ET D'AUTOMATISM 11	Zagreb 1,424
BIOSPECTRA d.o.o. 6	Rijeka 124
Bureau van Dijk Editions Electroniques SA 3	Split 113
Prysmian Netherlands BV 3	Osijek 80
Ultra Electronics Forensic Technology Inc 3	Velika 61

#### Companies: 7,658

Columns Click to Select-

Profile Link	Name	City	Country
BR	DJD ZAGREB d.o.o.		
BR	Anton Paar Croatia d.o.o.		
8 <sub>m</sub>	SUHOMONT d.o.o.		
BR	SPEGRA INĹ ˝ENJERING d.o.o.		
8 <sub>m</sub>	Aspecta		

## **View specific contractors**



## Search for contracting authorities

#### CROATIA

HOME DASHBOARDS SEARCH DOWNLOAD ABOUT DONATE

TENDER COMPANY AUTHORITY

#### Search Authority

#### Filter Click to Select -

Authority: Name			Authority: City				
Authority: Name	×	Authority: City	×	Authority: Main Activities	×	Authority: Buyer Type	×
Grad Zagreb	496	Zagreb	641	General Public Services	2,687 🔺	Public Body	2,881
Klinički bolnički centar Rijeka	186	Rijeka	163	Other	1,963	Regional Authority	2,630
Klinički bolnički centar Zagreb	185	Pula	128	Health	1,431	Other	1,946
<ul> <li>HŽ Infrastruktura d.o.o.</li> </ul>	151	Split	125	Education	841	Utilities	1,064
<ul> <li>Ministarstvo unutarnjih poslova</li> </ul>	116	Zadar	120	Water	457	National Authority	641
				Housing And Community A.	. 427	National Agency	57
				Electricity	320	Regional Agency	35
				Environment	240 -		

#### Authorities: 9,480

Columns Click to Select\*

rofile	Name	City	Main Activities	Buyer Type
link		ony		buyer type
	Lokalna razvojna agencija-PoduzetniÄŤki centar GareĹ nica d.o.o.	GareĹĭnica	Other	Other
	Općinski sud u Krapini	Krapina	General Public Services	Other
	Pučko otvoreno učilište Rab	Rab	Recreation Culture And Religion	Other
	Osnovna škola Pećine	RIJEKA		
	OPĆINA LOKVIČIĆI	KRIVODOL	General Public Services	Other

## View specific contracting authorities



ى طالە

#### Download the data

CROATIA

HOME DASHBOARDS SEARCH DOWNLOAD ABOUT DONATE

#### Download

On this page you find all the tender data of this portal. Data description.

Croatia	CSV	JSON	NDJSON
Tenders: 241582	Digiwhist 30.0 MB	Digiwhist 192.5 MB     OCDS 53.4 MB	Digiwhist 192.5 MB

All content (including data) on this website is licensed under Creative Commons BY-NC-SA 4.0.

Other countries		③ CSV description	JSON description	OCDS description	
Country	Tenders	csv	JSON		NDJSON
All Data	46734888	Digiwhist 9.9 GB	Digiwhist 31.2 GB		Digiwhist 31.2 GB
Austria	95156	Digiwhist 18.8 MB	Digiwhist 66.5 MB	• OCDS 31.3 MB	Digiwhist 66.5 MB
Belgium	96256	Digiwhist 22.7 MB	• Digiwhist 118.2 MB	• OCDS 47.6 MB	• Digiwhist 118.3 MB
Bulgaria	162316	• Digiwhist 61.9 MB	Digiwhist 333.6 MB	• OCDS 168.4 MB	Digiwhist 333.7 MB
Croatia	241582	• Digiwhist 30.0 MB	• Digiwhist 192.5 MB	• OCDS 53.4 MB	Digiwhist 192.5 MB
Cyprus	8944	Digiwhist 2.3 MB	• Digiwhist 9.8 MB	• OCDS 3.5 MB	Digiwhist 9.8 MB

## Thank you!

## Questions?

nfL

Ц Ц Iceland Liechtenstein Norway Norway grants grants

#### An overview of procurement integrity and introduction to opentender.eu

## **COVID specific risks example: North Macedonia**

Péter Horn

Second annual summer school of the R2G4P, 8 June 2022

Implemented by:



The R2G4P project, coordinated by the Center for the Study of Democracy, Bulgaria benefits from a € 1.5 million grant from Iceland, Liechtenstein and Norway through the EEA and Norway Grants Fund for Regional Cooperation. The aim of the project is to implement shared anti-corruption and good governance solutions in Southeast Europe through innovative practices and public-private partnerships.

### Total procurement spending on COVID related goods

- In North Macedonia the total value of COVID-related purchases between 2020q1 and 2021q2 was 1.6 times the aggregate value a year and a half before the pandemic
- The aggregate value was about the same as the total value between 2017q1 and 2019q4



# Corruption Risk Index of the North Macedonian healthcare sector

- The CRI of the North Macedonian COVID product market and the broader healthcare market have increased significantly
- However the CRI growth had already begun a quarter before the pandemic, therefore it is hard to separate the effects of the pandemic from other policy changes
- This highlights the importance of qualitative research besides quantitative analysis



## Biggest winners of extra funds in the healthcare sector

- The figure shows the quantile distribution of the additional funds received by firms that have won more funds between 2020q1 and 2021q2 compared to 2018q3 and 2019q4
- The top quantile (67 firms) received 94.62% of the extra funds
- The top 5 biggest winners (of additional funds) received 46% of the extra funds (EUR 404.7 mil)



## Thank you!

## Questions?

nfL

Ц Ц